DOCUMENT RESUME

ED 135 829 TM 006 054

AUTHOR Dirlam, David K.; Manganal, Richard A.
TITLE Structured Key-Word Analysis.
PUB DATE [Apr 76]
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Classification; Computer Oriented Programs; *Response Style (Tests); Scoring; *Structural Analysis; Student Testing; *Word Frequency
IDENTIFIERS *Key Word Analysis; Open End Questions

ABSTRACT

Structured Key-word Analysis is a computerizable technique for objectively analyzing the natural language responses to open-end questions. Key-words are frequently used content words. Pairs of key-words that appear together in protocols either extremely often or extremely seldom (compared with the probability of co-appearance) are termed conjoint or disjoint pairs respectively. Individual protocols can then be scored for degree of stereotyping, uniqueness and fluency. The analysis should be especially helpful for evaluating open education, since standardized tests with closed-end items may not be appropriate for that use. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED135829

Structured Key-word Analysis

DAVID K. DIRLAM and RICHARD A. MANGANAL
State University of New York at Plattsburgh

paper to be presented at the
American Educational Research
Association Annual Meeting,
April 1976

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM006 054

Structured Key-word Analysis

DAVID K. DIRLAM and RICHARD A. MANGANEL
State University of New York at Plattsburgh

Structured Key-word Analysis is a computerizable technique for objectively analyzing the natural language responses to open-end questions. Key-words are frequently used content words. Pairs of key-words that appear together in protocols either extremely often or extremely seldom (compared with the probability of co-appearance) are termed conjoint or disjoint pairs respectively. Individual protocols can then be scored for degree of stereotyping, uniqueness and fluency. The analysis should be especially helpful for evaluating open education, since standardized tests with closed-end items may not be appropriate for that use.

(a) Objective

The major objective of this research was to develop an objective, computerizable technique for reliably analyzing responses to open-end questions.

(b) Theoretical Framework

There is a dearth of objective techniques for analyzing the natural-language responses to open-end questions. For the last several years, one of the authors has been developing a theory of conceptual organization that suggests one means of doing such an analysis (Dirlam, 1972, 1975; Dirlam, Courtney, Uttich and Hays, 1974; Dirlam and Optiz, 1974). Basically, concepts are organized according to a "classificational structure": a generalization of the distinctions between addition and multiplication of classes made by Inhelder and Piaget (1964) and between hierarchies and spaces made by Posner (1973). Recently, we have developed a technique which employs word frequencies to generate a common classificational structure for answers to questions in a specifiable context by a particular group of subjects. This technique called "Structured Key-word Analysis" employs content words that are used by significantly more than one subject. These are referred to as "key-words." A basic assumption is that if two key-words commonly appear either together or not at all in the answers of different subjects, then regardless of whether they appear in the same sentence, they have a close structural relation for the particular context in question.

(c) Method

Data to be subjected to analysis consists of N_L lists of single-word verbal responses ($l_1, l_2 \ldots l_N$). Duplicate words in each list are deleted to produce a series of dictionaries ($d_1, d_2 \ldots d_N$). The dictionaries are combined to produce L, the master list. Duplicate words are deleted from L to produce D, the master dictionary. A key-word dictionary, KW, with N_{KW} words, is produced by accepting only those items from D which appear in a minimum of six of the original lists. The frequency of usage of a word within 25 or more lists must be at least six for it to be considered significantly more frequent than one usage.

For each word in KW a search is made through the N_L dictionaries to determine exactly in which of them that word appears. This is indicated by an ordered row of ones and zeroes with ones indicating that the word appeared in those dictionaries. This process is repeated for each word in KW, generating an occurrence matrix M with N_{KW} rows and N_L columns.

All pairwise combinations of the rows of M are examined to determine the following data: (1) the frequency of the first key-word (the number of different dictionaries or columns in which the first word appears); (2) the frequency of the second key-word (the number of different dictionaries in which the second word appears); and (3) the frequency of overlap (the number of different dictionaries in which both words appear simultaneously). The triplet thus produced is used with N_L to determine the probability, according to the cumulative hypergeometric law, that the overlap would be as large (or small) as it was due to chance. If the probability is very small, then the common usage of the pair of words is considered either conjoint or disjoint, depending on whether there is exceptionally much or exceptionally little overlap.

(d) Data source

The data most suitable to this type of analysis are 50 - 250 word responses in specifiable contexts from 25 to 50 different subjects. Before the final analysis, the function words are deleted (function words are those small, very frequent words in the language that have primarily a syntactic function; they are the words, for example, that are normally omitted from telegrams). Of the remaining words (called "content words") the nouns are converted to singular form and the verbs to infinitive form.

(e) Results

The method of analysis was employed in two problem situations to date (the Dirlam et al., 1974, study used a less well-developed form of the technique to analyze responses to a creativity questionnaire): (1) descriptions of self of fifth grade slow learners who have been homogeneously grouped by ability for several years and of those who have been heterogeneously grouped and (2) descriptions of self by normal seventh grade children.[1] In the first case it was found that there was a significantly larger number of words and smaller proportion of key-words used by the heterogeneously grouped slow learners than by those who were homogeneously grouped. From this information it is possible to conclude that the homogeneously grouped children developed more stereotyped self concepts. In the second case it was possible to develop a common classificational structure of the self-descriptions of the seventh graders that will be useful in determining whether certain procedures commonly used in humanistic education contribute to the development of self-knowledge.

(f) Educational and scientific importance of the study

A major application of this technique concerns the current problem of evaluating the performance of children in open education. Many professionals involved in open education object to the closed-end format of most standardized tests. Because children in this type of education are encouraged to understand traditional subject matter in an individualistic manner, it seems inappropriate to make them conform to answer choices that are completely specified in advance. Nevertheless, many other professionals have difficulty accepting the subjectivity that usually accompanies the analysis of responses to open-end questions. Even for such apparently simple judgments as scoring for uniqueness on a creativity questionnaire, there are difficulties. For example, are the following two uses of a knife identical: "A pendulum bob for making a building," and "hanging it on a string to see how steep a cliff is"? In the context of several such answers from each of a group of subjects, Structured Key-word Analysis would generate the same answer for all scorers.

[1]Jill Otis and Michael Carrino were collaborators with the authors on the first and second studies, respectively.

REFERENCES

Dirlam, D. K. Most efficient chunk sizes. Cognitive Psychology, 1972, 3, 335-359.

Dirlam, D. K. Developmental and educational applications of structure theory: A mathematical reformulation of Piaget's concept of structure. Paper presented at the fifth special invitational interdisciplinary seminar on Piagetian theory and its implications for the helping professions, Los Angeles, January, 1975.

Dirlam, D. K. and Opitz, D. The structural analysis of creativity. I. The mathematics of structures. Paper read at the Annual conference of the Wisconsin Association of Sciences, Arts and Letters. Green Bay, April, 1974.

Dirlam, D. K. The structural analysis of creativity. II. Empirical results. Paper read at the Annual Conference of the Wisconsin Association of Sciences, Arts and Letters. Green Bay, April 1974.

Inhelder, B. and Piaget, J. The early growth of logic in the child: Classification and seriation. New York: Norton, 1964 (tr.).

Posner, M. I. Cognition: An introduction. Glenview, Illinois: Scott, Foresman, 1973.